R E P O R T R E S U M E S

ED 018 157 43 FL 000 594

SOME REMARKS ON STIMULUS-RESPONSE THEORIES OF LANGUAGE LEARNING. PSYCHOLOGY SERIES, TECHNICAL REPORT NUMBER 97.

BY- SUPPES, PATRICK CROTHERS, EDWARD
STANFORD UNIV., CALIF. INST. FOR MATH. STUDIES

REPORT NUMBER BR-5-1088 PUB DATE 12 JUN 66
REPORT NUMBER TR-97
CONTRACT OEC-6-14-000
EDRS PRICE MF-$0.25 HC-$2.36 57P.

DESCRIPTORS- *LANGUAGE RESEARCH, *LEARNING THEORIES, *VERBAL OPERANT CONDITIONING, *PSYCHOLINGUISTICS, *SECOND LANGUAGE LEARNING, CONDITIONED RESPONSE, CONDITIONED STIMULUS, EXPERIMENTAL PROGRAMS, LANGUAGE INSTRUCTION, LINGUISTIC THEORY, RESEARCH PROJECTS, RUSSIAN, MATHEMATICAL APPLICATIONS, STIMULUS BEHAVIOR, BEHAVIOR THEORIES, BEHAVIORAL SCIENCE RESEARCH, ACOUSTIC PHONETICS, STANFORD UNIVERSITY, CALIFORNIA,

THIS TECHNICAL REPORT PRESENTS, IN SUBSTANCE, THE CONTENTS OF THE FIRST CHAPTER OF A BOOK IN PREPARATION DEALING WITH STANFORD UNIVERSITY EXPERIMENTS ON PARTICULAR ASPECTS OF THE LEARNING OF RUSSIAN BY AMERICANS WITH NO PREVIOUS EXPOSURE TO SLAVIC LANGUAGES. THE GENERAL AIM OF THE COMPLETED VOLUME WILL BE TO APPLY CERTAIN PRINCIPLES AND RESULTS OF MODERN LEARNING THEORY TO THE STUDY OF SECOND LANGUAGES BY YOUNG ADULTS. IN DELINEATING A CONCEPTION OF THE RELATIONSHIP BETWEEN PSYCHOLOGY AND LINGUISTICS, THE INITIAL CHAPTER DISCUSSES PSYCHOLOGY AND SECOND LANGUAGE INSTRUCTION, PSYCHOLOGICAL THEORY, AND LINGUISTIC THEORY AND SECOND LANGUAGE LEARNING. ALSO PRESENTED ARE SOME REMARKS ON THE BASIC THEORETICAL RESULTS FROM THE MATHEMATICALLY-BASED THEORIES OF CONDITIONING TO BE APPLIED IN THE SUCCEEDING CHAPTERS ON THE ACTUAL EXPERIMENTATION. A LIST OF PERTINENT REFERENCES IS APPENDED. (AB)

BR-5-1088
PA. 48

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED018157

SOME REMARKS ON STIMULUS-RESPONSE THEORIES OF LANGUAGE LEARNING

BY

PATRICK SUPPES AND EDWARD CROTHERS

TECHNICAL REPORT NO. 97

JUNE 12, 1966

PSYCHOLOGY SERIES

FL 000 594

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

TECHNICAL REPORTS

PSYCHOLOGY SERIES

INSTITUTE FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES

(Place of publication shown in parentheses; if published title is different from title of Technical Report, this is also shown in parentheses.)

1 D. Davidson, S. Siegel, and P. Suppes. Some experiments and related theory on the measurement of utility and subjective probability. August 15, 1955. (Experimental test of the basic model, Chapter 2 in Decision-making: An Experimental Approach. Stanford Univ. Press, 1957)

2 P. Suppes. Note on computing all optimal solutions of [illegible]al linear programming problem. November 15, 1955.

3 D. Davidson and P. Suppes. Experimental measurement of utility by use of a linear programming model. April 2, 1956. (Experimental test of a linear programming model, Chapter 3 in Decision-making: An Experimental Approach. Stanford Univ. Press, 1957)

4 E. W. Adams and R. Fagot. A model of riskless choice. August 7, 1956. (Behavioral Science, 1959, 4, 1-10)

5 R. C. Atkinson. A comparison of three models for a Humphreys-type conditioning situation. November 20, 1956.

6 D. Scott and P. Suppes. Foundational aspects of theories of measurement. April 1, 1957. (J. Symbolic Logic, 1958, 23, 113-128)

7 M. Gerlach. Interval measurement of subjective magnitudes with subliminal differences. April 17, 1957.

8 R. C. Atkinson and P. Suppes. An analysis of two-person game situations in terms of statistical learning theory. April 25, 1957. (J. exp. Psychol., 1958, 55, 369-378)

9 R. C. Atkinson and P. Suppes. An analysis of a two-person interaction situation in terms of a Markov process. May 29, 1957. (In R. R. Bush and W. K. Estes (Eds.), Studies in Mathematical Learning Theory. Stanford Univ. Press, 1959. Pp. 65-75)

10 J. Popper and R. C. Atkinson. Discrimination learning in a verbal conditioning situation. July 15, 1957. (J. exp. Psychol., 1958, 56, 21-26)

11 P. Suppes and K. Walsh. A non-linear model for the experimental measurement of utility. August 21, 1956. (Behavioral Science, 1959, 4, 204-211)

12 E. Adams and S. Messick. An axiomatization of Thurstone's successive intervals and paired comparisons scaling models. September 9, 1957. (An axiomatic formulation and generalization of successive intervals scaling, Psychometrika, 1958, 23, 355-368)

13 R. Fagot. An ordered metric model of individual choice behavior. September 12, 1957. (A model for ordered metric scaling by comparison of intervals. Psychometrika, 1959, 24, 157-168)

14 H. Royden, P. Suppes, and K. Walsh. A model for the experimental measurement of the utility of gambling. September 25, 1957. (Behavioral Science, 1959, 4, 11-18)

15 P. Suppes. Two formal models for moral principles. November 1, 1957.

16 W. K. Estes and P. Suppes. Foundations of statistical learning theory, I. The linear model for simple learning. November 20, 1957. (Foundations of linear models. In R. R. Bush and W. K. Estes (Eds.), Studies in Mathematical Learning Theory. Stanford Univ. Press, 1959. Pp. 137-179)

17 D. Davidson and J. Marshak. Experimental tests of a stochastic decision theory. July 25, 1958. (In C. W. Churchman and P. Ratoosh (Eds.), Measurement: Definition and Theories. New York: Wiley, 1959. Pp. 233-269)

18 J. Lamperti and P. Suppes. Chains of infinite order and their application to learning theory. October 15, 1958. (Pacific Journal of Mathematics, 1959, 9, 739-754)

19 P. Suppes. A linear learning model for a continuum of responses. October 18, 1958. (In R. R. Bush and W. K. Estes (Eds.), Studies in Mathematical Learning Theory. Stanford Univ. Press, 1959. Pp. 400-414)

20 P. Suppes. Measurement, empirical meaningfulness and three-valued logic. December 29, 1958. (In C. West Churchman and P. Ratoosh (Eds.), Measurement: Definition and Theories. New York: Wiley, 1959. Pp. 129-143)

21 P. Suppes and R. C. Atkinson. Markov learning models for multiperson situations, I. The theory. February 20, 1959. (Chapter 1 in Markov Learning Models for Multiperson Interaction. Stanford Univ. Press, 1960)

22 J. Lamperti and P. Suppes. Some asymptotic properties of Luce's beta learning model. April 24, 1959. (Psychometrika, 1960, 25, 233-241)

23 P. Suppes. Behavioristic foundations of utility. July 27, 1959. (Econometrica, 1961, 29, 186-202)

24 P. Suppes and F. Krasne. Application of stimulus sampling theory to situations involving social pressure. September 10, 1959. (Psychol. Rev., 1961, 68, 46-59)

25 P. Suppes. Stimulus sampling theory for a continuum of responses. September 11, 1959. (In K. Arrow, S. Karlin, and P. Suppes (Eds.), Mathematical Methods in the Social Sciences. Stanford Univ. Press, 1960. Pp. 348-365)

26 W. K. Estes and P. Suppes. Foundations of statistical learning theory, II. The stimulus sampling model. October 22, 1959.

27 P. Suppes and R. C. Atkinson. Markov learning models for multiperson situations, II. Methods of analysis. December 28, 1959. (Chapter 2 in Markov Learning Models for Multiperson Interactions. Stanford Univ. Press, 1960)

28 R. C. Atkinson. The use of models in experimental psychology. May 24, 1960. (Synthese, 1960, 12, 162-171)

29 R. C. Atkinson. A generalization of stimulus sampling theory. June 14, 1960. (Psychometrika, 1961, 26, 281-290)

30 P. Suppes and J. M. Carlsmith. Experimental analysis of a duopoly situation from the standpoint of mathematical learning theory. June 17, 1960. (International Economic Review, 1962, 3, 1-19)

31 G. Bower. Properties of the one-element model as applied to paired-associate learning. June 29, 1960. (Application of a model to paired-associate learning, Psychometrika, 1961, 26, 255-280)

32 J. H. Blau. The combining of classes condition in learning theory. August 23, 1960. (See Transformation of probabilities, Proceedings of the Amer. Math. Soc., 1961, 12, 511-518)

33 P. Suppes. A comparison of the meaning and uses of models in mathematics and the empirical sciences. August 25, 1960. (Synthese, 1960, 12, 287-301)

34 P. Suppes and J. Zinnes. Stochastic learning theories for a response continuum with non-determinate reinforcement. October 25, 1960. (Psychometrika, 1961, 26, 373-390)

35 P. Suppes and R. Ginsberg. Application of a stimulus sampling model to children's concept formation of binary numbers, with and without an overt correction response. December 14, 1960. (Application of a stimulus sampling model to children's concept formation with and without an overt correction response, Journal exp. Psychol, 1962, 63, 330-336)

(Continued on inside back cover)

SOME REMARKS ON STIMULUS-RESPONSE THEORIES OF LANGUAGE LEARNING

by

Patrick Suppes and Edward Crothers

TECHNICAL REPORT NO. 97

June 12, 1966

PSYCHOLOGY SERIES

Reproduction in Whole or in Part is Permitted for

any Purpose of the United States Government

INSTITUTE FOR MATHEMATICAL STUIDES IN THE SOCIAL SCIENCES

STANFORD UNIVERSITY

STANFORD, CALIFORNIA

Table of Contents

		Page
1	Introduction	1
1.1	Psychology and second-language instruction	3
1.2	Psychological theory	7
1.3	Linguistic theory and second-language learning	13
1.4	Some remarks on theories of conditioning	27
	Table	48
	References	49
	Footnotes	51

This Technical Report presents, in substance, the contents of Chapter 1 of a forthcoming book on some second-language learning experiments, by E. Crothers and P. Suppes (Academic Press).

SOME REMARKS ON STIMULUS-RESPONSE THEORIES OF LANGUAGE LEARNING[1]

Patrick Suppes and Edward Crothers

STANFORD UNIVERSITY

In broad outline, the aim of this book is to apply certain principles and results of modern learning theory to the study of second-language learning by young adults. In order to have a concentrated series of studies on a single language, all the studies reported in this book are concerned with Russian, and all the subjects of the experiments are speakers of native-American English, with no prior knowledge of a Slavic language.

This initial chapter delineates our conception of the relation between psychology and linguistics and presents, at least in elementary form, the basic theoretical results from mathematical learning theory that we apply in the remainder of the book. Each of the remaining chapters of the book reports several experiments concerned with a particular aspect of second-language learning. Chapter 2 describes studies on learning to discriminate auditorily presented Russian consonant- and vowel phonemes. In the experiments of Chapter 3, subjects hear a Russian word and are to learn its orthographic representation in the Cyrillic alphabet. Chapter 4 is devoted to vocabulary learning experiments, in which subjects learn the Russian "equivalents" of English words. Chapter 5 presents an analysis of

selected topics in the learning of noun and verb inflections from visually presented material. The topic of Chapter 6 is the learning of grammar by induction from auditorily presented Russian utterances. Finally, in Chapter 7 we indulge in a few speculations as to those directions for future research which appear profitable in the light of the findings reported in Chapters 2-6.

Within each chapter our objectives are three-fold: to collect empirical evidence on the roles of selected experimental variables, to specify how the rate of learning an item depends on its linguistic structure, and to formulate and test learning models for individual experiments. The organization within each chapter reflects these three interests: the results section of each experiment has separate subheadings on effects of experimental treatments, analyses of item difficulty, and applications of models. Hence the reader who wishes to bypass one or another of these aspects may readily do so. Also, the relative emphasis on these objectives varies from one chapter to another. For example, mathematical models are analyzed in detail in Chapters 2, 4, and 5, whereas they receive little attention in Chapters 3 and 6. The experiments are reported in practically the same order that they were originally conducted, and the chapter-to-chapter progression of topics reflects our changing interests and our desire to survey a wide range of experimental topics rather than to focus exclusively on a single facet of second-language learning. Perhaps it is more than an idle hope to think that the progression

also indicates that our ideas and interests were maturing a little. At any rate, our own bias is that all three objectives are pursued with more originality in Chapters 4-6 than in Chapters 2 and 3. On this matter the reader may form his own judgment, because the chapters are largely independent of one another(except for an occasional reference to a model that was introduced earlier). Finally, it is very important to remark that the use of complex, natural-language stimuli renders a certain amount of tedious detail inevitable in the description of materials and results. To avoid submerging the major points in the morass of detail, we usually preface the extended description by an overview of the experimental design. In addition, relegation of inessential details to separate appendices has made it easier to highlight the main development.

1.1 Psychology and second-language instruction

It is a truism sometimes obscured in the heat of current debate, that linguistics as it is now conceived does not tell us how to organize the materials of a second language for initial learning. In principle, psychological learning theory should be able to provide the guide lines for such organization. It is also quite clear that we cannot proceed from general systematic principles of learning theory to the details of such organization. In one sense, the inability to do so represents a failure of contemporary psychology. On the other hand, it should be apparent that the scientific task of proceeding from general principles to the detailed organization of

language teaching is exceedingly complex, certainly much more so than any problems yet solved in linguistics or psychology. In order to clarify this point, let us consider a few examples of the kind of decisions that are needed. What vocabulary size (e.g., 20, 30, or 50 items) should be employed during the initial hours of instruction in Russian? In principle, there should be an application of mathematical learning theory that provides an optimal result. But even granted that this question can be answered, we still have not resolved the more pressing problem of exactly which items (words) should be introduced. Should we select the words in some simple fashion from a frequency count of word occurrences in spoken Russian? Or should we begin primarily with a few nouns and some verbs reflecting the regular first conjugation? In this same vein, a more global problem is to specify the relative proportions of time allotted to phonology, vocabulary, and grammar training. Since our ultimate objective is mastery of the language, and not merely of vocabulary, we may reformulate our earlier question about initial vocabulary size. That is, should training on word inflections be introduced early (in which case we will restrict ourselves to a modest-sized vocabulary), or should it be postponed (in which case the initial training may be vocabulary drill, with a larger list)? Practical decisions along this line must be made by every teacher of Russian, and corresponding questions arise in the teaching of any other second language. It is also apparent that, as yet, systematic principles for making these

decisions are very far from being available.

We would like to be able to offer an empirically verified prescription for solving these questions in the teaching of Russian or any other foreign language. Unfortunately, we are not able to specify such principles. Nor do we expect to discover them in the immediate future. In this book we do hope to contribute an accumulation of scientific results on particular aspects of learning Russian as a second language. Our results are incomplete, in at least two important respects. First, our decision to conduct detailed analyses of selected aspects made it unfeasible to examine every aspect. The most noteworthy example here is that pronunciation learning was not analyzed in its own right (although it was investigated in conjunction with grammar learning). Second, a particular subject participated in only one experiment; we have not yet attempted to integrate the various aspects into a single long-term instructional routine. The main reasons for relegating each aspect to an isolated experiment stemmed from our interest in applying mathematical models. It would be uneconomical to run an extended experimental course when the model made predictions for only one segment of the course. Also, a theoretical analysis of learning in the later segments would be complicated by transfer effects. Additionally, a practical limitation that should be mentioned at the outset was that all of our Russian speech stimuli were recorded by the same native speaker. As to how our findings on these individual aspects can be fitted into the

classroom practices for teaching Russian, we must leave the decisions to the teacher and textbook writer. The pedagogic implication of the research is that it places increasingly stronger constraints on teachers and textbook writers. From the standpoint of qualitative results of the sort described in this book and generally available in the psychological literature, it would not be difficult to write a fairly devastating critique of most of the introductory textbooks in Russian. This, however, is not our purpose in this book. Our intention is to contribute to the constructive literature by reporting the results of carefully controlled experimentation on topics running from phoneme discrimination to the learning of grammar rules. For example, the teacher who wants to know what Russian phoneme discriminations are difficult for native Americans can consult the data reported in Chapter 2. For the teacher or writer who wants to know about certain problems of vocabulary acquisition, we believe that the experiments reported in Chapter 4 provide useful information.

In this connection we should remark that, until recently, people not engaged in psychological research have been inclined to belittle possible practical applications of such research. With respect to second-language learning, one reason was that many of the early experiments by educational psychologists were plagued by poor experimental design. Although studies in the area of verbal learning were more carefully controlled, they were usually limited to the learning of verbal paired associates and the like, using English or nonsense

material. While these investigations have led to the discovery of significant variables for verbal learning, the relevance to second-language learning may be remote, owing to obvious profound differences between stimuli from a foreign natural language and these verbal-learning stimuli. Because of this same question of relevance, we rejected the use of artificial-language material, preferring instead to use miniature systems consisting of authentic Russian phonemes, words, and sentences. Indeed, as other experimenters have found, it is not easy to isolate the pedagogically significant variables even when one is using authentic second-language material. Many of the variables which we expected to produce marked effects had either no effects or unanticipated effects. While we have been able to illuminate the roles of a number of variables, it will come as no surprise that many other variables remain to be explored.

1.2 Psychological theory

The learning theory that we apply in this book is a variant of stimulus-sampling theory, which was sketched in its present form in a fundamental paper by Estes (1950). In the broader context of psychological theories, this theory is essentially a stimulus-response theory. In view of the controversy that surrounds the applicability and adequacy of stimulus-response theories for language learning, some general remarks seem necessary in this introductory chapter. These comments are intended to guard against misunderstanding in appreciating the range and limitations of the claims we make for the application of

theory to detailed experiments, such as the investigations reported in later chapters.

The first important point is this. We do not claim that stimulus-sampling theory in its current formulation is sufficiently complex or rich enough in structure to provide a detailed understanding of language learning. This is an inescapable criticism of stimulus-sampling theory, but what is to be emphasized once this point is accepted is that we would make the same claim about any other theory either in psychology or linguistics. No existing psychological or linguistic theory can account for any substantial portion of the systematic details of language learning. No doubt psychologists who have written in stimulus-response frameworks have usually over-estimated the power of their theory and underestimated the complexities of language learning. Because of the considerable discussion - on the part of both psychologists and linguists - about the adequacy of psychological and linguistic theory, it will perhaps be useful for us to expand upon these remarks in some detail.

We shall first give an informal axiomatic characterization of stimulus-sampling theory and then discuss its adequacy for the facts of language learning. Models of this theory will be applied in later chapters, but in the present context we shall be concerned more with general ideas than with detailed elaboration of particular

models. The axiomatic formulation given here follows that of Suppes and Atkinson (1960). The axioms are expressed verbally, but it is reasonably clear how they may be converted into a formulation that is mathematically rigorous within the framework of modern probability theory. The axioms depend upon four basic concepts of stimulus-response psychology; namely, stimulus, response, reinforcement, and conditioning, plus the concept of stimulus sampling. Essentially, the theory conceptualizes the sequence of events that takes place on a trial as follows. A set of stimuli are presented to the organism. From this set the organism samples a single hypothetical stimulus element or stimulus pattern. He then responds, and the actual response made depends on the current conditioning state of the sampled element. After the response is made a reinforcing event occurs and, depending upon the nature of the reinforcing event, the conditioning of the sampling stimulus is or is not changed. States of conditioning are postulated, and the reconditioning of the sampled stimulus places the organism in a new state. The sequence of events then is repeated on the next trial. The occurrences of the various events described are governed by probability laws, as is made clear in the statement of the axioms below. Readers unfamiliar with contemporary psychological theory in its quantitative aspects might ask about certain kinds of

restrictions that occur in the statement of the axioms. For example, why are the axioms restricted to situations involving discrete trials? Why is it assumed that the subject samples only a single stimulus on each trial rather than a heterogenous set of stimuli? The answers to these queries are to be given partly in terms of mathematical convenience. The extensions to handle either continuous time or sampling of large sets of stimuli are conceptually straightforward but technically awkward. For reasons that will become clear subsequently in this discussion, we feel that the main difficulties of the theory are not centered around these restrictions, but around more fundamental conceptual issues.

The axioms as formulated are meant to apply to a finite set of stimuli, a finite set of responses,and a finite set of reinforcing events, with a natural 1-1 correspondence obtaining between responses and reinforcing events. The axioms are divided into three groups: the first group dealing with the sampling of stimuli, the second with the conditioning of sampled stimuli, and the third with responses.

Sampling axioms

S1. Exactly one stimulus element (pattern) is sampled on each trial.

S2. Given the set of stimulus elements available for sampling on a trial, the probability of sampling a given element is independent of the trial number and the preceding pattern of events.

Conditioning axioms

C1. On every trial each stimulus element is conditioned to at most

one response.

C2. If a stimulus element is sampled on a trial, it becomes conditioned with probability c to the response (if any) that is reinforced on that trial; if it is already conditioned to that response, it remains so.

C3. If no reinforcement occurs on a trial, there is no change in conditioning on that trial.

C4. Stimulus elements that are not sampled on a given trial do not change their conditioning on that trial.

C5. The probability c that a sampled stimulus element will be conditioned to a reinforced response is independent of the trial number and the preceding pattern of events.

Response axioms

R1. If the stimulus element sampled on a trial is conditioned to a response, then that response is made.

R2. If the stimulus element sampled on a trial is not conditioned to any response, then one of the possible responses is made in terms of a guessing distribution that is independent of the trial number and the preceding pattern of events.

There are several things to be noted about these axioms. In the first place they seem to formulate the entire theory of information processing in terms of the conditioning of stimuli and not at all in terms of more explicit cognitive processes. But this distinction is more apparent than real. Vague talk about cognitive processes is

itself not very enlightening until a specific theory of cognitive processes is assumed. An interesting question then is what are the formal relations between models of stimulus-sampling theory as formulated here and models of the proposed cognitive theory. It is shown, for example in Suppes and Atkinson (1960), that for the application of certain cognitive theories to experiments in probability learning, there exists a formal isomorphism between models of stimulus-sampling theory and models of the proposed cognitive theory. By referring to this example, which is worked out in detail in the first chapter of Suppes and Atkinson, we do not mean to suggest that such a formal isomorphism can be found for all learning situations or all theories. What we do mean to suggest is that the relation between stimulus-sampling and conditioning ideas on the one hand and cognitive ideas on the other cannot be discussed in scientifically serious terms until the two corresponding theories are given a specific formulation. The thesis that we would want to defend about the apparent conflict between behavioristic and cognitive theories is that much of the conflict is apparent rather than real. When the theories are formulated in a mathematically sharp fashion and in terms that suffice to deal with the details of any substantial body of experimentation, then a surprising amount of agreement in formal structure is to be found, in spite of the rather different terminology that is used.

We would contend that the most striking thing about behavioral and cognitive theories of learning is that they mainly share the same

important weaknesses. All extant theories, or at least all the theories known to us, have as their central failure a lack of a structure which is rich enough to provide an account of the learning of any complex problems. To us, it is quite an indifferent matter as to which framework - cognitive or behavioristic - will ultimately prove most helpful in formulating this richer structure. Certainly there is a current tendency to use the cognitive language appropriate to computers in searching for notions suitable for the analysis of learning, and it may well turn out that this direction will be an important one for current research. Whether the language is behavioristic or cognitive in tone is of little importance, we feel, compared to the question of whether or not the theory has been formulated in a mathematically viable fashion. The history of psychology from Hume to Hull is strewn with theories that were stillborn from any reasonable mathematical viewpoint. We would maintain that until a theory is capable of clear mathematical expression it is scarcely a systematic theory at all.

1.3 Linguistic theory and second-language learning

In Sec. 1.1 we made some general remarks on the failure of current linguistic or psychological theories to provide an adequate account of second-language learning, and in Sec. 1.2 we discussed at greater length the stimulus-sampling learning theory that has formed the theoretical background of most of the experiments reported in this book. Now we consider, in a discursive way, some of the alleged

shortcomings of stimulus-response psychology as an approach to a theory of language learning. Our argument will not attempt to refute the criticisms; in fact, we agree with many of them. Therefore we shall not review these critiques in detail. Rather, our argument is that the critics have not offered a satisfactory replacement for the stimulus-response approach to language learning. The rest of this section is devoted to an amplification of this assertion.

In the last decade, linguists have eagerly seized upon these defects of psychological theories, and have enunciated a number of constructive criticisms. On the other hand, some linguists seem to feel that linguistic theory itself is able to offer a proto-psychological theory of language learning. In this section we shall review some representative claims of linguists, and say why these claims fail to inspire a more realistic account of language learning.

Before considering particular examples, it may be useful to indicate in a general way what we think are the main weaknesses of the viewpoint and methodology of linguists with respect to second-language learning. To a psychologist who reads the linguistic literature on these matters, undoubtedly the single most striking characteristic of linguists' pronouncements on language learning is the frequent indifference to presenting or analyzing any systematic empirical data. Whether the point under discussion is concerned with the learning of phonology, or of the morphemes of a given language, or of the generative rules of grammar of the language, the discussions usually rely on

impressionistic evidence. No empirical tests of generative grammars have been made, at least not in the detailed fashion that has characterized mathematical psychology during the past decade. Evidently this is because a theory of grammar is not itself a theory of performance, and at present any predictions of performance are based on somewhat hazardous extrapolation from the formal theory. If the predictions are not fulfilled, one can take refuge by repudiating the informal extrapolation, while still maintaining that the theory of grammar is correct. Until the gap between theory and linguistic performance has been bridged in a mathematically precise way, the theory is essentially untestable, which probably explains why the number of purportedly relevant experiments is small. Perhaps the second most striking characteristic of this linguistic literature is the contentious philosophical tone. Since most of the published writings are neither mainly concerned with systematic presentations of bodies of data, nor with formal logical and mathematical systems, it is not surprising that the viewpoint is strongly oriented towards philosophical methods of discourse. Of course, we do not mean to denigrate philosophical methods of discourse, but we do think that classical philosophical methods of reasoning are an insufficient and inappropriate approach to a subject that is inherently scientific and empirical in character.

The third general observation is the unusual degree to which linguists are concerned to provide counterexamples to show that psychological theories are incapable of handling the facts of

language learning. Our attitude needs to be stated with some care. It is certainly appropriate to provide counterexamples when psychologists assert exaggerated claims about the explanatory power of their theories. We do not want to attempt to defend the many kinds of statements made by psychologists about the adequacy of psychological theory to explain language learning. We would agree with the linguists that present-day theories are certainly inadequate to the task. However, it is well known that in virtually every area of active scientific investigation one can readily produce examples that cannot be handled by the current theory. It is just as easy to do this in physics as in psychology, but the cavalier production of such counterexamples cannot be regarded as a constructive step toward a more satisfactory theory, unless the counterexamples are accompanied by definite suggestions for modifying or replacing the theory.

Another point demands attention here. Many linguists have been most enthralled by what they call the theory of competence, which is the kind of theory that has been extant in mathematics for a very long time. Consequently they seem to believe that the theory of competence can be used on any occasion to demonstrate that a particular psychological approach is fruitless. To our mind, this indiscriminate use of the theory of competence is as misguided as continued refutation of Newtonian mechanics by referring to the theory or phenomena of color. Clearly Newtonian mechanics, as classically formulated, cannot give an account of the production and changes of color of objects over time corresponding to the prediction

of their trajectories of motion, but this does not invalidate the theory in some total fashion. Later we shall discuss in more detail the inappropriate use of the theory of competence.

A detailed analysis of all major linguistic comments on language learning and psychological theories of language behavior would be too serious a digression from the main purpose of the present book. Moreover, the overwhelming preponderance of this literature is directed to the enumeration of deficiencies in psychological theories of first-language learning. We could cite many publications in linguistics which dwell on problems of language learning, but which dismiss issues of second-language learning with the banal remark that everyone knows there are fundamental differences between first-language and second-language learning.

The theoretical reasons for concentrating on first-language learning are apparent, and seem to be justified. On the other hand, it is clear that from a pedagogical standpoint a better psycholinguistic theory about the learning of second languages would be a very desirable development. We would also surmise that as the theoretical literature on second-language learning develops,many of the schisms current between linguists and psychologists will re-emerge in the analysis of second-language learning. Let us, then, examine some of the issues more closely, and also attempt to ascertain their implications for second-language learning. It will suffice to confine our remarks to the viewpoints expressed in recent books by Chomsky (1965) and Katz

and Postal (1964), as well as the recent exchange between Bever, Fodor and Weksel (1965) on the one hand, and Braine (1965c) on the other.

An important feature of this literature is the pre-eminent role it assigns to the theory of competence. Roughly speaking, this theory is defined to be the theory of the language itself, apart from consideration of precisely how it is acquired and used by speakers and listeners. It is characteristic of these discussions to emphasize the primacy of the theory of competence even for the development of the theory of performance - the latter being the theory of actual language behavior.

Presumably the major goal of the theory of competence is to develop a theory of syntax, semantics, and phonology for a spoken natural language or class of languages. Being more amenable to attack, the problems of developing a theory of syntax have received far more attention than those of developing a theory of semantics, and for that reason most of our own remarks will be directed toward the former. However, insofar as learning and performance are concerned, it is our conviction that semantics may well turn out to be more important. Once a comprehensive and adequate theory of semantics of natural languages is developed, it will likely entail a major revision in conceptions of syntax. In succeeding chapters that report detailed experiments on second-language learning, the theory of competence will rarely be mentioned. Therefore it is appropriate now to attempt

to justify this omission, and to say why we think the importance of this theory for first - or second - language learning has been over-emphasized.

1. Assuming that the theory of competence furnishes an adequate syntax for the natural spoken language that is to be taught as a second language, we would like to make our first point by analogy to the study of mathematics learning. The formalization of mathematics within well-defined artificial languages has been for several decades an important part of investigations into the foundations of mathematics. In particular, once a given body of mathematics is formalized in such a language (that is, the formal language is stated, together with rules of inference and axioms of a non-logical sort, for the mathematics) then a large number of general questions about the body of mathematics in question can be precisely discussed. There are three examples that suggest analogies to problems of language learning. The first is that it is a simple matter in a formalized language to give a recursive definition of the well-formed formulas. As everyone recognizes, such definitions are incredibly simpler than the generative grammars that seem to be required for natural languages. But it still also seems true that for purposes of recognizing whether or not a particular expression is well formed, the formal recursive definition itself is seldom used by individuals who work with such a logical language. In difficult or doubtful cases,appeal to the formal definition will indeed take place; typically it will not. Instead, individuals seem to use certain explicitly organized heuristics as

cues of recognition.

A simple instance of this is the following. Consider the recursive definition of a well-formed formula in sentential logic.

a. 1. The single letters 'p'. 'q'. and 'r' with or without numerical subscripts are formulas.

b. 2. If S is a formula, then $\neg$(S) is a formula.

c. 3. If S and T are formulas, then (S)&(T), (S)$\vee$(T), (S)$\rightarrow$(T) and (S)$\leftrightarrow$(T) are formulas.

d. 4. A finite sequence of symbols of the language is a formula only if its being so follows from the above rules.

Now consider the expression

$$(((p)\rightarrow(q))\vee(r)\ \&\ \neg(s)(.$$

Even the novice does not have to apply the formal definition of a formula, working from the inside out and checking each step. Rather, he can instantly recognize that the expression is not a formula. Why? Because he will notice at once the left parenthesis at the right-hand end of the expression, and he need investigate no further. If people resort to heuristics even where the formal characterization is relatively simple, then a fortiori we would expect them to adopt strategies when confronted with a language having a complex generative grammar. Unfortunately, we do not have systematic empirical data on this question, and therefore shall not explore it further.

The second example, however, is well corroborated by general

experience and therefore is perhaps more appropriate. It concerns the matter of discovering formal proofs of theorems. In principle, it is quite straightforward to give an algorithm for all proofs. One simply begins by enumerating the proofs and eventually any proof will turn up in this list after only a finite number of predecessors. Thus, if a certain conjecture is proposed as a theorem one can begin to enumerate proofs, and if the conjecture is indeed a theorem at some point it will be produced as a proof. If the conjecture is not a proof then this procedure will not, of course, establish this conclusion. The point is, however, that any proof will be produced by this simple algorithmic procedure. But surely no one would seriously suggest this algorithm as a feasible method of proving theorems. The analogy to learning a language should not be pressed too far, but the basic point is valid; namely, that the existence of algorithms for finding proofs or of formal grammars for characterizing a natural language grammar hardly guarantees that subjects do in fact employ these particular algorithms or generative rules, or that the rules even have substantial relevance to the actual method of learning.

The third example may be cited to amplify this last remark. It concerns the relation between the theory of games and the actual learning to play a game skillfully. For a game of perfect information (e.g., chess) it can be proven that there is a pure strategy such that if a player adopts it, he is ensured of at least a tie in every game. The proof goes back to Zermelo (1912). And for a game of imperfect

information (e.g., bridge) we know from fundamental results of von Neumann (1928) and von Neumann and Morgenstern (1944) that optimal mixed strategies exist for each player. Moreover, the games mentioned are wholly finitistic, and in the case of bridge the total number of bids and plays is not inordinately large. But the complete enumeration of strategies for chess or bridge is far beyond the capabilities of even the best computers, and the analytical computation of optimal strategies is similarly impractical. How, then, do people actually learn to play chess or bridge? It is a question we cannot answer, but there do seem to be cogent reasons for thinking that the mathematical theory of games has little relevance to actual behavior in these more complicated games. Game theory and a theory of competence are analogous in the following sense: neither intends to consider limitations of human information-processing capacities, and neither intends to consider the mnemonics and strategies which people invent to utilize their capacities more effectively.

In this connection we offer two subsidiary remarks about the concept of infinity in a theory of competence. The first is to record our impression that linguists concerned with the theory of competence and with the fact that a generative grammar will generate an infinity of sentences are rather too impressed with this infinity of possibilities. For example, Bever, Fodor, and Weksel (1965, p. 481) propose as a serious criticism of Braine's work that "no language which consists of a finite set of strings requires phrase-structure rules

in its grammar, for any such language can be enumerated by a simple list". As Braine rightly remarks in his reply, this point is correct only if subjects do actually learn by enumerating. Finite lists of any substantial size are not learned in this rote fashion, and from the standpoint of language learning there is certainly no sharp distinction to be made between a collection of 10^{100} sentences and an infinite collection of sentences. The implication from Bever, Fodor, and Weksel's remark is that subjects would learn by an enumeration routine, simply because such a routine exists. But this supposition is unwarranted, for roughly the same reason that the existence of an algorithm for discovering proofs does not ensure that people employ the algorithm. If one is going to object to a finite language, the meaningful objection is not that phrase-structure rules are unnecessary. Rather, it is that the imposition of finite bounds creates mathematical difficulties in the recursive system. We shall return to this matter later, in discussing questions of probability measures on the lengths and compositions of sentences.

Secondly, we want to cite another analogy to express our skepticism that the theory of competence as now formulated will be of serious systematic help in developing an adequate theory of performance. This analogy derives from computer science. A decade or so ago many people fondly hoped that the theory of recursive functions as developed extensively in mathematic logic would be of major use in the foundations of computer theory. It is fair to say that this has

turned out not to be the case, and for reasons that seem apparent. The classical theory of recursive functions involves infinite domains and unbounded operations, whereas the theory of actual computers is necessarily restricted to bounded finite systems. There is good reason to believe that it is precisely the finitistic limitation of actual computers that is responsible for the lack of deeper application of the theory of recursive functions in computer science. Admittedly, we have a relatively clear understanding of the finitistic limitations of the computers now constructed, and we have a much less refined understanding of the finitistic limitations of human powers of learning and memory. Nonetheless, the existence of finite limitations to human capabilities is a fact too obvious to require demonstration. The importance of these finitistic restrictions is sufficient to provoke suspicion that the theory of competence may be irrelevant, just insofar as it does deal with an infinite collection of objects.

2. Our second general reason for neglecting the theory of competence in the chapters that follow is the absence of any probabilistic element in currently formulated theories of competence. We have already mentioned one simplifying abstraction of the theory of competence - that it admits sentences of arbitrary length. A case might be made for the admission of such sentences if at the same time the theory of competence were rich enough to derive the probability distributions on sentences. The simplest kind of marginal distribution might well be in terms of sentence length, and

here it is apparent that as the length of a sentence became arbitrarily large the probability measure assigned to sentences in this class would become arbitrarily small, for any reasonable theory. In order that this point not be misunderstood we emphasize the word marginal in the characterization of the distributions. We would hardly suggest that an adequate theory of competence that took into account the distributional character of sentences, phrases, morphemes, phonemes, etc. would regard sentence length as being fundamental. Certainly the assigned probability measure would be a function of sentence structure. Nevertheless, it would be odd indeed if the marginal distribution of sentence lengths was not essentially unimodal in character, with sentences of longer and longer length being assigned smaller and smaller probabilities. A theory of performance that included derivations of probability distributions for linguistic units in actual speech would probably be quite worthwhile from the standpoint of second-language learning. Certainly this information would permit an exacting test of the theory, and it might well suggest what sentence structures should be emphasized in language instruction.

From the standpoint of the application of mathematically formulated theories in psychology to the analysis of data from systematic experiments, it is fair to say that the most important methodological gain made in the past decade has been the realization that theories need to be formulated probabilistically in order to provide the proper degree of tightness in expressing the relation between theory and data. Human behavior as we now understand it, be it speech or any other variety

of behavior, is too complicated to expect that an algebraic theory will predict the major phenomena with reasonable accuracy. In this respect the theories of classical physics that served so long as a model of scientific theorizing have indeed turned out to be badly misleading.

The motivation for introducing probabilistic notions seems especially compelling if one concentrates on spoken language, unrectified by the well-defined conventions of the printer. Although it is mathematically convenient to ignore the complexities of actual speech while concentrating on a theory that is several steps removed from such actual speech, it must be acknowledged that this is a highly simplifying abstraction. It is especially this sort of abstraction that causes one to doubt that any algebraic theory of competence is directly relevant to the subtle facts of language learning.

Let us just give one simple example of important probabilistic considerations that have been excluded from theories of grammaticality but that are essential to a full-fledged theory of performance. These are the considerations surrounding variables of timing and speed, as exemplified by the response latency experiments in Chapter 4. As far as we know, no theory of competence takes into account timing variables in speech, and yet from the standpoint of comprehension it is an eminently critical variable, particularly for second-language learning. Almost anyone can acquire the rudiments of a second language fairly readily if that language is spoken very slowly (e.g., a word every 10 seconds) and with precise articulation. What is more

significant, however, is to study learning under conditions of normal speaking rate. For example, it would be our judgment that problems of timing are more crucial than problems of grammaticality in the initial phases of second-language learning. In work initiated since this book was written, we are concerned primarily with examining the effects of pacing variables on production and comprehension of a spoken second language. We hope in subsequent publications to be able to elaborate on this point, which we are presently making only in a superficial way. To reiterate the conclusion from the foregoing arguments, we believe that the idealized native speaker whom writers on the theory of competence like to conjure up should be modeled on a stochastic process and not along algebraic lines.

1.4 Some remarks on theories of conditioning

In view of the widespread use of stimulus-response theories of conditioning, it is natural that they are a favorite target of linguistic attacks. As should be clear from Sec. 1.2, stimulus-sampling theory as formulated there is one variant of conditioning theory. Thus it seems incumbent upon us to comment on the relation between such a theory and language learning, paying particular attention to those criticisms of stimulus-response theory that have been voiced in discussions of the theory of competence.

To repeat, we are not interested in making any last-ditch defense of the thesis that classical conditioning theory is sufficient for explaining the complexities of verbal behavior. Rather, we seek to

put into perspective some of the linguistic criticisms, and attempt to show why we think they are not as devastating as their authors claim. A representative criticism of conditioning theory is to be found at the end of a book by Katz and Postal (1964). This passage is the closing part of a two-page final section on implications of their book for the theory of language learning. (The P-markers referred to in the quotation are phrase markers.)

> Purely inductive abstraction from observable properties of phonetic objects in the child's corpus cannot, in principle, explain how the child learns to understand the meaning of sentences, because many of the syntactic features on which the meaning of sentoids depends are nonexistent in final derived P-markers and thus are in no way physically marked in phonetic objects. Hence, there are no observable features to indicate how a child can obtain a semantic interpretation that depends on information about syntactic properties not represented in final derived P-markers. But without such observable aspects of sentence structure from which to abstract, a conditioning theory has no basis for an abstraction that accounts for the way one relates semantic interpretations to phonetic objects. For any conditioning theory -- by definition -- presupposes observable aspects

> of a stimulus (in this case, aspects of sentence structure) to which something else (in this case, semantic features, however construed) is conditioned. Therefore, since no account of how children learn the meaning of sentences is possible without the formulation of this richer structure found in underlying P-markers, a conditioning theory of language acquisition must be rejected as being, in principle, incapable of explaining how language is learned.

The phrase that we want especially to comment on is the last one. "A conditioning theory of language acquisition must be rejected as being, in principle, incapable of explaining how language is learned." This passage appears to rest on a fundamental misunderstanding as to how stimulus-response theories are now being used in psychology. That is, it appears to make the unjustified assumption that stimulus-response psychology is bound by the very rigid restriction that all its theoretical constructs have immediately obvious observable counterparts. Later we shall examine this point in some detail.

First, however, we wish to voice our disagreement with another implication of the passage. It seems to suggest that the only theory worth developing is an ideal theory which will account for all the phenomena in question. But surely any proposed theoretical venture would be doomed by such a demanding standard, even Katz and

Postal's theory. To clarify this statement, consider the following two theses.

Thesis 1. Since no fully adequate account of the meaning of sentences is possible without the formulation of a theory about the formation and changes of beliefs held by the speakers and listeners of the sentences uttered, a semantic theory of the sort proposed by Katz and Postal must be rejected as being, in principle, incapable of an adequate formulation of semantics.

Thesis 2. Since no current generative grammar includes a real-time component that accurately predicts temporal properties of speech, any generative grammar as currently formulated must be rejected as being, in principle, incapable of explaining the actual grammatical structure of spoken language.

We think that these two theses are about as sound as the Katz and Postal claim about conditioning theory, but we do not at all propose that they are devastating criticisms of the interesting work in semantics by Katz and Postal, or the very substantial work in generative grammars that has been done by Chomsky, by Harris (1951), and by their collaborators in the past decade and a half. Instead, the role of the Katz and Postal criticism should be to stimulate new extensions of conditioning theories, just as Thesis 2, we believe, urges the inclusion of a stochastic element in generative grammars.

To avoid misunderstanding, we would like to state our point more precisely. First, we assume that linguists who criticize conditioning

theory for being too simple would like to support their contentions by an exact analysis. In effect, they would want to show that, given a mathematically sharp formulation of a psychological theory and a canonical formulation of accepted data about natural language or users of natural language, then it could be shown formally that the theory in question could not possibly explain the accepted data. We concur with Katz and Postal that conditioning theory as it now stands is inadequate in practice and can be proven inadequate in principle. More explicitly, we feel that there are sentences describing accepted data that cannot be derived as predicted results within any present-day theory of conditioning. At an even deeper level, we believe that there are concepts needed to describe agreed-upon data of language learning that cannot be defined in terms of the fundamental concepts of any extant theory of conditioning.

However, our point in the present discussion is to emphasize our belief (cf. Thesis 1) that this is true of any semantic theory now extant in relation to its explanation of the meaning of sentences, and also true of the grammar of a spoken language (cf. Thesis 2). Thus, we think that our two theses are in this respect just as sound as that of Katz and Postal. Our procedure is like theirs in that we are not offering systematic data and a rigorous analysis that precisely justifies the theses. But we believe that all three theses would generally be regarded as valid statements about ways in which current theoretical undertakings fall short of our ambitious standards for a truly comprehensive theory.

As noted earlier, Katz and Postal's criticism, and especially the phrase "in principle", appears to rest upon a very pessimistic appraisal of prospects for future growth and extension of conditioning theories. If their quotation were simply that any current conditioning theory of language is incapable of explaining how language is learned, there would be immediate general agreement among all but the most entrenched. The addition of the phrase "in principle" constitutes a very much stronger claim, and it is this stronger claim that we now want to examine more carefully. To begin with, we must confess that we do not fully understand exactly what is meant by "in principle". We shall attempt to present and analyze two possible explications of what the phrase "in principle" might conceivably be taken to mean.

1. A first meaning of "in principle" is that there is no conservative extension of the theory of conditioning which would explain major aspects of language learning. By "conservative" we mean that the extension would employ only the same fundamental concepts as the original theory.

An example is the following. It is well-known that the three classical problems of squaring the circle, trisecting any angle, and doubling a cube cannot be solved by means of straightedge and compass construction alone. Moreover, it is possible to give a precise axiomatization of plane geometry in terms of constructive concepts, and to show that the models of these axioms are just those isomorphic to a two-dimensional vector space over a Euclidean field. (A Euclidean

field is an ordered field that contains the square root of every positive element.) However, by using existential quantifiers, but without changing the constructive concepts of the theory, it is possible to add axioms that yield an extension of the theory, and moreover have the property that any models of the extended theory are just the standard ones of two-dimensional vector spaces over the field of real numbers. Of course within the framework of this extended theory, the three classical problems are solvable. In this geometric example we have a precise specification of what the original and extended theories can do, and especially of what extensions are admissible. It is just this precision that is totally missing from the Katz and Postal discussion, and the absence renders ambiguous their usage of "in principle".

2. A second and much stronger meaning of "in principle" is that there is no extension of the theory of conditioning, even with addition of new fundamental concepts, which can explain language learning. We doubt that Katz and Postal intended this meaning, because such a claim seems outrageously strong. Therefore perhaps the first, weaker, meaning of "in principle" above is closer to the one they intended. If so, their claim would certainly be easier to defend. But it would be a compromise, and no longer an unqualified assertion that it is futile to develop stimulus-response theories of language learning. About the only hope of establishing anything in terms of the stronger meaning of "in principle" would be to establish that the theory of conditioning is logically complete. However, for reasons to be indicated now,

we feel that the theory is actually very incomplete, and that this very incompleteness enables one to adapt and extend the theory to areas which at first glance appear to lie beyond its scope.

In order to be more definite in the ensuing discussion, we shall refer to stimulus-sampling theory as formulated in Sec. 1.2, and not attempt to make remarks applicable to every theory of conditioning that may be found in the literature over the past decade or two. We agree wholeheartedly with many of Chomsky's (1959[a])criticisms of Skinner's (1957) claims about the ability of his version of conditioning theory to explain the facts of language learning. We also disavow any claim that stimulus-sampling theory provides a substitute theory able to substantiate Skinner's extravagant claims. On the other hand, we do consider it important to indicate in a general way our estimate of the hopes and prospects of stimulus-sampling theory for playing a significant role in some future theory of language learning. It will be apparent that most of our remarks in this connection apply both to first-language and second-language learning; this is not because we think the two processes are identical, but because at this stage of investigation any theory proposed for either process suffers from many of the same fundamental deficiencies.

There are two senses of incompleteness which apply to stimulus-sampling theory. One is the standard logical sense mentioned earlier in connection with the theory of conditioning. From a mathematical standpoint it is clear that the theory formulated in Sec. 1.2 is not complete, because it certainly does have essential extensions. We

would conjecture that future progress toward completing the theory will involve, in an important way, additional assumptions about stimulus complexity and stimulus structure. Obviously not only language learning but every form of complex learning and perception requires a more elaborate conception of stimulus structure. For example, an adequate account of visual perception could hardly be derived within the framework of stimulus-sampling theory unless much of the geometry of perception were somehow included in the theory. In succeeding chapters we make a number of detailed remarks about stimulus structure. Many of these remarks are not theoretical formulations of stimulus structure, but merely experimental analyses of how learning varies from one kind of item to another. In those instances where we actually have been able to express specific stimulus variables within a model, the model has been applicable to only one or two kinds of experiments. Thus, unhappily, we have no single unified theory that explicates particular structural variables over a wide range of experiments. Despite this limitation, we feel that the separate theoretical ventures have increased our understanding of language learning, at least of the second-language learning of Russian. At the same time, the cumulated body of experimental evidence helps us to identify exactly which variables are responsible for most of the variance in the data. Knowing this, we are more likely to include important variables, rather than trivial ones, in any future theory. For example, the vocabulary experiments of Chapter 4 show that learning depends more on properties of the Russian member of the

paired-associate than on the English member. Another interesting example is reported in Chapter 6 on grammar learning. Acquisition of Russian grammar is found to be influenced more by the availability of English translations than by either the presentation order or the particular words used to exhibit the grammar.

The second sense of incompleteness of stimulus-sampling theory concerns the multiplicity of possible empirical interpretations of what is meant by "stimulus", "response", and "reinforcement". We shall not dwell here on the notion of "reinforcement", because many of the comments to be made about "stimulus" apply equally to "reinforcement". By and large, the elementary event of reinforcement has been mainly characterized in the psychological literature as a 0, 1 event, or at most an event varying in intensity on a scale of preference. For complex experiments, reinforcement should be conceptualized in terms of what information it conveys to the subject. As stated, it suffices to limit our comments to stimuli, because in our experiments whenever two items differed in their post-response reinforcements they usually differed also in their pre-response stimuli. (An exception was our investigations of the role of redundant relevant auditory information when the visual information is logically sufficient to learn the language skill in question; pertinent research is reported in Chapters 3 and 5). Thus most of the important problems of interpretation can be reduced to questions about how the stimulus should be characterized. As we shall see, the notion of stimulus in stimulus-sampling theory is conspicuously incomplete, and hence so is the entire

theory.

Customarily, there is a fairly clear experimental interpretation of what events are to be classified as responses and as reinforcements, so that the canonical form of the observed data specifies in a well-defined discrete fashion the responses and reinforcements occurring on every trial. In the more general case when time is treated as a continuous rather than a discrete parameter, the responses and reinforcements are still treated as observable. The situation is radically different regarding the stimuli postulated to be present in the experiment. There are no established rules of correspondence between the hypothetical stimulus elements and the physical stimuli, so neither the stimulus population nor the stimulus sample can be identified unequivocally. Everyone agrees that it would be highly desirable ultimately to have such correspondence rules. But because each of the presently proposed rules lacks general applicability, the degrees of freedom available for contriving new rules are welcomed by theorists, and regarded as an essential strength of stimulus-sampling theory. The strategy of treating the stimulus as an unobservable entity, then, provides at the present time just about the right degree of slack in applications of the theory. As many people have recognized, it is just when a theory has all of its fundamental concepts formulated directly in terms of observables that it fails to fit data; the power of theoretical abstraction is unwisely forfeited by the insistence on strict experimental identifiability.

It is important to emphasize the difference between stimulus structure and stimulus identifiability. A richer characterization of structure seems essential to any account of more complex learning; on the other hand, it does not seem wise to insist that the hypothetical stimulus elements be directly identified in terms of observable stimulus properties.

Because we have not resolved the critical matters of stimulus structure here, and because we have been unable to construct an adequate general theory in subsequent chapters, we conclude this chapter with an example of how the problem might be approached. The example pertains to the phoneme-discrimination experiments to be reported in Chapter 2. Even though the stimulus structure in these experiments is quite simple compared to that in syntax- and morphology-learning experiments, the example is useful in several ways. One is that it indicates how the sampling axioms S1 and S2 of Sec. 1.2 can be related to assumptions about structure. Another is that it makes more concrete the problems of satisfactorily conceptualizing structure, and simultaneously emphasizes that the issues will not be resolved by any facile shift from the behavioristic language of conditioning to the mentalistic language of cognition.

The task we shall consider is that of learning to discriminate between Russian voiced and unvoiced consonants in pairs of consonant-vowel (CV) syllables. From the standpoint of distinctive features analysis, the phonemic contrasts involved are minimal. But from a more detailed psychological standpoint a number of variables enter

the picture, and their effects are not easily specified. For simplicity of analysis we shall restrict ourselves to the initial consonants /p/ and /b/,[2] presented auditorily to the subject. In the task we have in mind, the subject is asked to judge whether a CV:CV pair he hears represents the same or different consonants. For example, if the pair happens to be /pu:pu/, he should say "same", whereas if it is /pu:bu/ he should say "different". The vowel is always the same in both members of a given pair. To avoid additional complications, we shall omit considerations that revolve around stimulus-timing parameters, although a theory would certainly be incomplete unless it included an account of how learning depends on the durations of the various events and inter-event intervals.

The first step in the analysis of stimulus structure for this discrimination task is to characterize more exactly the set S of stimuli. For purposes of this example, we shall use the distinctive-features analysis of Halle (1959), and postulate a subset of stimuli for each distinctive feature. The primary eleven he lists are: vocalic, consonantal, diffuse, conpact, low tonality, strident, nasal, continuant, voiced, sharped and accented. For discrimination of a single phoneme we could postulate that S is simply the union of these eleven subsets. The example being considered here is considerably more complex, but before turning to it, there is a point about axioms S1 and S2 that may be made in connection with the simple task of recognizing a single phoneme(in order to make our theoretical point we ignore the questionable realism of trying to sound single phonemes).

Suppose single phonemes are sounded and the subject responds by printing or typing a phonemic symbol to represent graphemically what he thinks he heard. Under the most obvious sort of assumptions the subject samples various distinctive features of the phoneme--of course, not necessarily all of those present. According to the sort of conditioning theory described in Sec. 1.2, the sampled stimuli become conditioned to the correct response--shown to the subject by a correction procedure when he makes an error. When the subject samples a subset of S he makes a given response according to the proportion of sampled stimuli conditioned to that response. Note that this assumption is not the same as axioms S1 and S2. The difficulty of the theory presented in Sec. 1.2 is that it implies that subjects could never learn to discriminate perfectly the various phonemes. This prediction follows because the phonemes overlap in their distinctive features. For example, suppose that the stimulus phoneme were /p/ and consonantal and low-tonality stimulus elements were sampled and conditioned to the correct graphemic response. Then if /b/ were the next stimulus phoneme, there would be a positive probability of an incorrect response; the subject would sometimes write "p" instead of "b".. This error has positive probability, because at least some of the consonantal and low-tonality stimulus elements were conditioned to the grapheme "p" on the previous trial. In fact, under the above-stated assumption, the error probability would remain positive even after any finite number of reinforced trials.

Sampling axioms S1 and S2 are intended to circumvent this

difficulty. Within mathematical psychology, they are a first departure from atomistic views of stimulus structure, views that had their roots in the British associationist tradition of Hume and J. S. Mill. What is postulated by S1 and S2 is that the subject samples a pattern of stimulus elements, rather than a subset of elements individually conditioned. One formal way of defining these patterns is to transform S into the Cartesian product of the eleven subsets, or more simply for the present purpose, into a set of ordered 11-tuples. The i^{th} member of a tuple is a member of the i^{th} distinctive-feature subset. Or, if the feature is absent, the i^{th} member is the empty set 0. Thus /p/ would be represented by <0, c, 0, 0, t, 0, 0, 0, 0, 0, 0>, where c is a consonantal feature and t a low-tonality feature. (We emphasize that 0 here is the empty set and does not have the special meaning of Halle's 0 which designates a nonphonemic feature.) For purposes of simplicity we shall not introduce any principles of generalization across phonemes, although such postulates would seem essential to any complete analysis. Hence we simply apply S1 and S2 directly. The subject samples exactly one 11-tuple, i.e., one pattern, on each trial. The fundamental difference is that he responds according to the conditioning of the pattern, not according to that of individual stimulus elements.

The basic idea of this major extension of conditioning theory was first clearly enunciated by Estes (1959). It is clear, however, that the notion of patterns cannot immediately be extended to the recognition of larger linguistic units, for it would require that

each new utterance be treated as a new pattern which is as yet unconditioned to any response. To overcome the dilemma, we need some theoretical principle whereby different presentations can be treated as instances of the same pattern. As to what the principle should be, no facile general answer is possible, because any answer to the question of what the subject perceives as a unit is highly dependent on the overall stimulus situation. However, the problem is less severe for the present special case of phoneme discrimination, where it seems reasonable to treat each phoneme as a pattern. Doing so does not beg the question of phoneme identification, because of the well-known psychological distinction between perceiving something as a unit (i.e., as a pattern) and identifying it.

As we have mentioned, in Experiments I and II of Chapter 2 the subject was confronted with a contrast between a voiceless- and a voiced consonant phoneme in a pair of CV syllables. What sort of model might capture the essentials of the discrimination process? A major requirement for any prospective model is that it be able to predict which contrasts will be easy and which ones will be difficult. To make matters more concrete, let us consider the /b:p/ contrast when the vowel is /a/. Four kinds of CV pairs exemplify this contrast: they are /ba:ba/, /pa:pa/, /pa:ba/, and /ba:pa/. Of course, the correct answer is "same" for each of the first two pairs, and "different" for each of the last two. We have listed these pairs in ascending order of difficulty, as measured by the proportions of errors obtained in the experiments to be reported in the next chapter.

There is reason to think that this rank order reflects something fundamental to the discrimination process, because the same order was found with all other vowels and stop consonants. If we let U and V denote an unvoiced CV syllable and a voiced CV syllable, respectively, then invariably the empirical rank order from least to most difficult was /V:V/, /U:U/, /U:V/, and /V:U/. Clearly, it is not sufficient for a model merely to reproduce this rank order. It should also be able to give a reasonably accurate prediction of the proportion of errors on each type of CV pair. The model to be discussed does meet these requirements.

The rank order did not change as a function of the number of learning trials, so in the model we shall ignore learning and attempt to reproduce the rank order. It would be a fairly easy matter to attach a simple learning mechanism to the model, because the only important condition is that the mechanism not allow the rank order to be a function of the trial number. However, consideration of learning would only introduce an unnecessary complication.

To characterize the model, we extend the basic theory of Sec. 1.2 in the following way. We suppose that to attempt the desired comparison the subject samples a pattern from the first CV, stores it in a memory register, samples a pattern from the second CV, and then makes a comparison. At what stages does failure of this mechanism generate errors? There are two rather natural ways to proceed. One is to postulate a decay function for the storage of the first CV of each

temporally ordered pair. The other is to postulate a sampling failure, or, in other words, an attention failure, in hearing the second CV. In the present case, the latter of these two sorts of postulates explains the observed data much better than does the former. When a sampling failure does occur, we postulate a guessing probability distribution over the two possible responses, which is the sort of assumption used with considerable success in many recent learning studies such as Atkinson and Crothers (1964), Bernbach (1965), Millward (1964b), and Suppes, Groen, and Schlag-Rey (1966), and is already embodied in axiom R2 of Sec. 1.2. Formally, we extend the theory of Sec. 1.2 by assuming the following special sampling axiom for this experimental situation.

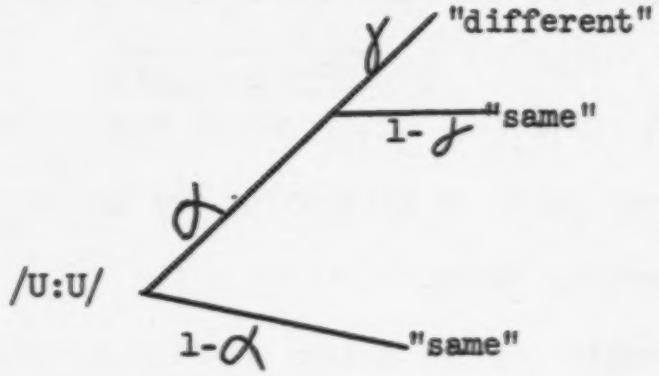
S3. With probability α, a voiceless second syllable is not sampled as a pattern, and with probability β a voiced second syllable is not sampled as a pattern.

As a merely technical extension of response axiom R2 we postulate:

R2'. If no pattern is sampled from the second CV, then one of the possible responses is made in terms of a guessing distribution that is independent of the trial number and the preceding pattern of events.

Naturally we would prefer to give a more direct phonological rationale of α and β, but we see little hope of doing so in the near future. It does seem reasonable to attach different parameters to the voiced and voiceless consonants.

The derivation from S3, R2' and the other axioms of Sec. 1.2 of the probability of an error on each type of CV-pair is straightforward. First, to obtain a mathematical expression of R2', let γ be the guessing probability of responding "different" and therefore $1-\gamma$ the probability of responding "same" . Then the probability, P(U:U), of an error on a /U:U/ pair is just the probability $\alpha\gamma$ of not sampling as a pattern the second U and then making the wrong guessing response. A simple tree diagram show the possibilities.



By similar argument we compute the probability of an incorrect response upon presentation of each of the other three types of pairs. These quantities are:

$$P(U:U) = \alpha\gamma$$
$$P(V:V) = \beta\gamma$$
$$P(V:U) = \alpha(1-\gamma)$$
$$P(U:V) = \beta(1-\gamma)$$

According to the data reported in Chapter 2 the corresponding observed error proportion early in learning, based on data from all vowels, were .16, .07, .45 and .21 for the /p:b/ contrast. Estimating α, β and γ from these data, we obtain $\hat{\alpha} = .61$, $\hat{\beta} = .28$ and $\hat{\gamma} = .26$, which yield predictions exactly accurate to two (but not to three)

decimal places. Recognizing that it is not optimal simply to carry over this estimate of the guessing probability γ to the other two contrasts /t:d/ and /k:g/, but in order to give an impression of what may be done in a simple way with the model formulated, we may retain the estimate $\hat{\gamma}$ = .26 and proceed as follows. By adding P(U:U) and P(V:U), we get an estimate of α, and by adding P(V:V) and P(U:V), we get an estimate of β for /t:d/ and for /k:g/. The results are quite satisfactory; they are summarized in Table 1.1. In fact, the

Table 1.1 here

/p:b/ predictions are slightly better for the /t:d/ contrast than are the predictions based on estimating two parameters, because the observed proportions are so close.

The extension of the axioms of Sec. 1.2 has been rather modest in the present case. For an exact mathematical treatment we would need to specify more exactly the definition of a trial in order to make the interpretation of axioms S1 and S3 completely clear. For example, it is implicit in the extension described here that we treat the sampling of the first CV pair as one "trial" and the sampling of the second as a second trial, even though no overt response is required between the drawing of the two samples. In a more general treatment we would proceed along the lines of Suppes and Donio (1965) and treat time as a continuous rather than as a discrete parameter.

However, it is clear to us, and we are sure it is clear to our readers, that the fundamental conceptual problem that we have not yet touched is to extend the theory of Sec. 1.2 to the central linguistic

phenomena of understanding and speaking meaningful sentences. Until that is done, even if only in rough approximation, it cannot be claimed that a satisfactory theory of language learning has been formulated. We do not know what form such a theory will take. We do think it will be surprising if the conditioning mechanisms that are central to stimulus-response theories do not play an essential part. What we are not yet able to do is to formulate the additional structural constraints required for complex language behavior. The aim of this book is to explore some of the directions that may permit at least some progress on these difficult problems, and at the same time to present the empirical results of a large number of systematic experiments, which in themselves impose serious constraints on any future theory.

Table 1.1

Proportions of errors in discriminating Russian voiced:voiceless stops

	/p:b/		/t:d/		/k:g/	
	obs.	pred.	obs.	pred.	obs.	pred.
P(U:U)	.16	.16	.14	.16	.06	.09
P(V:V)	.07	.07	.07	.08	.04	.04
P(V:U)	.45	.45	.46	.44	.27	.24
P(U:V)	.21	.21	.22	.21	.10	.10
$\hat{\alpha}$	-	.61	-	.60	-	.33
$\hat{\beta}$	-	.28	-	.29	-	.14
$\hat{\gamma}$	-	.26	-	.26	-	.26

References

Atkinson, R. C., and Crothers, E. J. A comparison of paired-associate learning models having different learning and retention axioms. J. math. Psychol., 1964, 1, 285-315.

Bernbach, H. A. A forgetting model for paired-associate learning. J. math. Psychol., 1965, 2, 128-144.

Bever, T., Fodor, J. A., and Weksel, W. On the acquisition of syntax: A critique of "contextual generalization".. Psychol. Rev., 1965, 72, 493-500.

Braine, M. D. S. On the basis of phrase structure: a reply to Bever, Fodor, and Weksel. Psychol. Rev., 1965c, 72, 483-492.

Chomsky, N. Review of: B. F. Skinner, Verbal behavior. Lang., 1959a, 35, 26-58.

Chomsky, N. Aspects of the theory of syntax. Cambridge: M. I. T. Press, 1965.

Estes, W. K. Toward a statistical theory of learning. Psychol. Rev., 1950, 57, 94-107.

Estes, W. K. Component and pattern models with Markovian interpretations. In R. R. Bush and W. K. Estes (Eds.), Studies in mathematical learning theory. Stanford: Stanford Univer. Press, 1959, 9-52.

Halle, M. The sound pattern of Russian. Gravenhage: Mouton, 1959.

Harris, F. S. Methods in structural linguistics. Chicago: Univer. Chicago Press, 1951.

Katz, J. J., and Postal, P. M. An integrated theory of linguistic descriptions. Cambridge: M. I. T. Press, 1964.

References (cont.)

Millward R. An all-or-none model for noncorrection routines with elimination of incorrect responses. J. math. Psychol., 1964b, 1, 392-404.

Skinner, B. F. Verbal behavior. New York: Appleton-Century, 1957.

Suppes, P., and Atkinson, R. C. Markov learning models for multiperson interactions. Stanford: Stanford Univer. Press, 1960.

Suppes, P., and Donio, J. Foundations of stimulus-sampling theory for continuous-processes. Tech. Rep. no. 69, Institute for Mathematical Studies in the Social Sciences, Stanford Univer., 1965.

Suppes, P., Groen, G., and Schlag-Rey, M. A model for latencies in paired-associate learning. J. math. Psychol., 1966, 3, 99-128.

von Neumann, J. Zur theorie der Gesellschaftsspiele. Mathematische Annalen, 1928, 100, 295-320.

von Neumann, J., and Morgenstern, O. Theory of games and economic behavior. Princeton: Princeton Univer. Press, 1944 (2d ed., 1947).

Zermelo, E. Über eine Anwendung der Mengenlehre auf ide Theorie des Schachspiels. Proc. of the Fifth International Congress of Mathematicians. Cambridge, II, 1912, 501-510.

Footnotes

[1] This project was supported by a grant from the Carnegie Corporation of New York and by the U. S. Office of Education (OE 6-14-009).

[2] Slanted lines denote phonemes.